On the Geometry Of Acceptability Functionals

Alois Pichler[a,*]

[a] *University of Vienna, Austria. Department of Statistics and Operations Research.*

Abstract

In this paper we discuss the geometry of acceptability functionals or risk measures. The dependence of the random variable is investigated first. The main contribution and focus of this paper is to study how acceptability functionals vary whenever the underlying probability measure is perturbed.

It turns out that the Wasserstein distance provides a valuable notion of distance, and may acceptability functionals allow a precise quantification in terms of this distance.

Keywords: Risk Measures, Continuity properties, Wasserstein and Kantorovich Distance
2010 MSC: Optimal transport,

1. Introduction

In this paper we discuss evaluations of acceptability functionals (risk measures) for different probability measures. A major motivation to study this problem is given by the fact that applications and optimization problems, which arise in finance or energy, use acceptability functionals in their objective. Moreover decisions on optimal investment, which are often driven by (stochastic) optimization and which involve acceptability functionals, frequently build on empirical distributions. But similar observations are possibly as likely as the observations at hand, so it is desirable to ensure that the acceptability functionals employed will result in some similar values as well for similar observations.

It turns out that the Wasserstein distance (Kantorovich distance) is a useful notion of distance which ensures the desired properties: Various acceptability functionals allow an estimation in terms of the Wasserstein distance, and for important ones these estimates can be given in a sufficiently precise form.

Above that we describe – for some selected acceptability functionals – probability measures which are worst in the sense that they modify the resulting acceptability utmost.

Major investigations in the present paper are based on the Legendre-Fenchel transform and on Kusuoka's representation of acceptability functionals .

The paper is organized as follows: A preliminary discussion introduces the concepts and tools required. Section 3 contains a main result, an upper bound of potential results of acceptability functionals, whenever the measure is being perturbed. Section 4 characterizes measures which change the acceptability functional utmost, and particular evaluations are collected in Section 5. Some illustrations in Section 6 complete the paper.

2. Preliminary Discussion

Definition 1. A function $\mathcal{A}\colon \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ defined on a linear space of $\mathbb{R}-$valued random variables $\mathcal{Y}$ is said to be an *acceptability functional* if the following axioms are satisfied (compare the notion of *coherent risk measures* in [ADEH99]):

(i) MONOTONICITY: If $Y_1 \leq Y_2$, then $\mathcal{A}(Y_1) \leq \mathcal{A}(Y_2)$ [1] $(Y_1, Y_2 \in \mathcal{Y})$;

*Corresponding Author
URL: `http://isor.univie.ac.at/` (Alois Pichler)

[1] For Y_1 and Y_2 random variables we write $Y_1 \leq Y_2$ whenever $Y_1 \leq Y_2$ almost everywhere.

(ii) Concavity: $\mathcal{A}(tY_1+(1-t)Y_2) \geq t\mathcal{A}(Y_1)+(1-t)\mathcal{A}(Y_2)$ whenever $0 \leq t \leq 1$ $(Y_1, Y_2 \in \mathcal{Y})$;
(iii) Translation Equivariance: If $c \in \mathbb{R}$, then $\mathcal{A}(Y+c) = \mathcal{A}(Y)+c$; [2]
(iv) Positive Homogeneity: For $t>0$, $\mathcal{A}(tY) = t\mathcal{A}(Y)$ $(Y \in \mathcal{Y})$.

In this definition we allow $\mathcal{A}$ to evaluate to $-\infty$. However, throughout this paper we shall assume that $\mathcal{A}$ is *proper*, that is to say there is at least one Y with $\mathcal{A}(Y) > -\infty$.

Examples of Acceptability Functionals

Average Value-at-Risk

The most prominent acceptability functional probably is the *Average Value-at-Risk at level* α, which is defined as

$$\mathrm{AV@R}_\alpha(Y) := \frac{1}{\alpha}\int_0^\alpha \mathrm{V@R}_p(Y)\,\mathrm{d}p, \tag{1}$$

where $\mathrm{V@R}_p(Y) = \inf\{t\colon P(Y \leq t) \geq p\}$ is the left side quantile. The equivalent expression

$$\mathrm{AV@R}_\alpha(Y) = \max_{q\in\mathbb{R}} q - \frac{1}{\alpha}\mathbb{E}(q-Y)^+ \tag{2}$$

was elaborated in [RU00].

Distortion Acceptability Functionals and Kusuoka Representation

The Average Value-at-Risk, by (1), assigns the same weight to any of the quantiles which arise with probability p less than α. One may assign different weights, which gives rise to the following definition.

Definition 2 (Distortion acceptability functional and Kusuoka representation)**.** Let $\mathcal{A}\colon \mathcal{Y} \to \mathbb{R}\cup\{-\infty\}$ be an acceptability functional.

(i) $\mathcal{A}$ is a *distortion acceptability functional* provided that

$$\mathcal{A}(Y) = \int_0^1 \mathrm{V@R}_p(Y)\,h(p)\,\mathrm{d}p \tag{3}$$

with h satisfying

(a) $h \geq 0$ (to ensure monotonicity),
(b) $1 = \int_0^1 h(p)\,\mathrm{d}p$ (to ensure translation equivariance) and
(c) h decreasing (to ensure concavity);

(ii) A representation

$$\mathcal{A}(Y) = \inf_{m\in\mathscr{M}} \int_0^1 \mathrm{AV@R}_p(Y)\,m(\mathrm{d}p), \tag{4}$$

where $\mathscr{M}$ is a set of probability measures on $[0,1]$, is called *Kusuoka representation*.

A comprehensive discussion of distortion acceptability functionals was given in [Pfl06].

Law invariant Acceptability Functionals

All acceptability functionals introduced above are already determined by the law (the cumulative distribution function) of the considered random variable Y, for which they are called *law invariant*, or sometimes *version independent*.

It was elaborated by Kusuoka ([Kus01, JST06]) that every upper semi-continuous and law invariant acceptability functional on L^∞ has a representation as in (4).

[2] The random variable $Y+c$ is $Y + c\cdot\mathbb{1}$ where $\mathbb{1}$ is the constant random variable, $\mathbb{1}(x) = 1$.

Lemma 3. *Let $\mathcal{A}_h$ be a distortion acceptability functional. Then there is a probability measure m_h on $[0,1]$ such that*

$$\mathcal{A}_h(Y)=\int_0^1 \mathrm{V@R}_p(Y)\,h(p)\,\mathrm{d}p=\int_0^1 \mathrm{AV@R}_\alpha(Y)\,m_h(\mathrm{d}\alpha). \tag{5}$$

Conversely, for any probability measure m on $[0,1]$ there is a function h such that (5) holds for $\mathcal{A}_h$, provided that $\int_0^1 \frac{1}{\alpha} m(\mathrm{d}\alpha)<\infty$.

Moreover for every acceptability functional $\mathcal{A}$ there is a set $\mathscr{H}$ of bounded distortion functions such that $\mathcal{A}(Y)=\inf_{h\in\mathscr{H}}\mathcal{A}_h(Y)$.

Proof. One may associate with the density h of a distortion acceptability functional $\mathcal{A}_h$ the measure $m_h(A):=h(1)\,\delta_1(A)-\int_A \alpha\mathrm{d}h(\alpha)$ [3] on $[0,1]$. This is a positive measure, as h is decreasing and whence $-\int_A \alpha \mathrm{d}h(\alpha)\geq 0$. And it is a probability measure, as

$$m_h([0,1])=h(1)-\int_0^1 p\mathrm{d}h(p)=h(1)-h(1)+0\cdot h(0)+\int_0^1 h(p)\,\mathrm{d}p=1$$

by integration by parts.

With this choice

$$\begin{aligned}\mathcal{A}_h(Y)=&\int_0^1 \mathrm{V@R}_p(Y)\,h(p)\,\mathrm{d}p=\\ &=h(1)\int_0^1 \mathrm{V@R}_p(Y)\,\mathrm{d}p-\int_0^1 \mathrm{V@R}_p(Y)\,(h(1)-h(p))\,\mathrm{d}p\\ &=h(1)\,\mathrm{AV@R}_1(Y)-\int_0^1 \mathrm{V@R}_p(Y)\int_p^1 \mathrm{d}h(\alpha)\,\mathrm{d}p\\ &=h(1)\,\mathrm{AV@R}_1(Y)-\int_0^1\int_0^\alpha \mathrm{V@R}_p(Y)\,\mathrm{d}p\mathrm{d}h(\alpha)\\ &=h(1)\,\mathrm{AV@R}_1(Y)-\int_0^1 \mathrm{AV@R}_\alpha(Y)\,\alpha\mathrm{d}h(\alpha)=\int_0^1 \mathrm{AV@R}_\alpha(Y)\,m_h(\mathrm{d}\alpha),\end{aligned}$$

any distortion acceptability functional $\mathcal{A}_h$ thus is a combination of AV@Rs at different risk levels α.

As for the converse relation consider $\mathcal{A}_m(Y):=\int_0^1 \mathrm{AV@R}_\alpha(Y)\,m(\mathrm{d}\alpha)$ for some probability measure m on $[0,1]$. The function $h_m(p):=\int_p^1 \frac{1}{\alpha}m(\mathrm{d}\alpha)$ is positive, decreasing, and

$$\int_0^1 h_m(p)\,\mathrm{d}p=\int_0^1\int_p^1 \frac{1}{\alpha}m(\mathrm{d}\alpha)\,\mathrm{d}p=\int_0^1\int_0^\alpha \frac{1}{\alpha}\mathrm{d}p\,m(\mathrm{d}\alpha)=\int_0^1 m(\mathrm{d}\alpha)=1.$$

Employing (1) again gives

$$\begin{aligned}\mathcal{A}_m(Y)\;&=\;\int_0^1 \mathrm{AV@R}_\alpha(Y)\,m(\mathrm{d}\alpha)=\int_0^1\frac{1}{\alpha}\int_0^\alpha \mathrm{V@R}_p(Y)\,\mathrm{d}p\,m(\mathrm{d}\alpha)\\ &=\;\int_0^1 \mathrm{V@R}_p(Y)\int_p^1\frac{1}{\alpha}\,m(\mathrm{d}\alpha)\,\mathrm{d}p=\int_0^1 \mathrm{V@R}_p(Y)\,h_m(p)\,\mathrm{d}p,\end{aligned}$$

such that $\mathcal{A}_m$ is a distortion acceptability functional.

For the latter statement consider a probability measure m on $[0,1]$ and define the measures $m_n(A):=m\left(\left[0,\frac{1}{n}\right[\right)\cdot\delta_{\frac{1}{n}}(A)+m\left(A\cap\left[\frac{1}{n},1\right]\right)$. As $\alpha\mapsto\mathrm{AV@R}_\alpha(Y)$ is increasing it follows that $\int_0^1 \mathrm{AV@R}_\alpha(Y)\,m(\mathrm{d}\alpha)\leq\int_0^1 \mathrm{AV@R}_\alpha(Y)\,m_n(\mathrm{d}\alpha)$, and as $\alpha\mapsto\mathrm{AV@R}_\alpha(Y)$ is continuous and bounded

[3]Without loss of generality one may require $h(1)=0$, then $m_h(A)=-\int_A \alpha\mathrm{d}h(\alpha)$.

moreover that $\int_0^1 \text{AV@R}_\alpha(Y)\, m_n(\mathrm{d}\alpha) \to \int_0^1 \text{AV@R}_\alpha(Y)\, m(\mathrm{d}\alpha)$ whenever n tends to infinity. Note next that $\int_0^1 \frac{1}{p} m_n(\mathrm{d}p) \le n \cdot m\left(\left[0, \frac{1}{n}\right[\right) + n \cdot \int_{\frac{1}{n}}^1 m(\mathrm{d}p) = n < \infty$. It follows that $h_{m;n}(\alpha) := \int_\alpha^1 \frac{1}{p} m_n(\mathrm{d}p) \le n$ is bounded (by n) and $\mathcal{A}(Y) = \inf_n \mathcal{A}_{h_n}(Y)$. Employing Kusuoka's representation (4) and collecting all distortion functions as $\mathscr{H} = \{h_{m;n} \colon m \in \mathscr{M},\, n \in \mathbb{N}\}$ reveals the assertion. □

2.1. Continuity of the Acceptability Functional

In the sequel we shall restrict the investigations to the Banach lattice $\mathcal{Y} = L^p(X, \mathcal{F}, P)$ where $1 \le p \le \infty$. The axioms imposed on acceptability functionals have strong regularizing properties. For future reference in this paper we repeat here some of the key properties of an acceptability functional with fixed probability measure.

Monotonicity, together with translation equivariance in the definition of the acceptability functional, force $\mathcal{A}$ to be Lipschitz-continuous on the subspace L^∞, as the following lemma reveals.

Lemma 4. *Suppose that $\mathcal{Y} = L^\infty(X, \mathcal{F}, P)$ and there is one almost surely bounded random variable $\tilde{Y} \in L^\infty$ such that $\mathcal{A}\left(\tilde{Y}\right) > -\infty$. Then $\mathcal{A}$ is finite valued ($\mathcal{A}(Y) > -\infty$)* for any $Y \in L^\infty$ *and $\mathcal{A}$ has Lipschitz constant 1, that is*

$$|\mathcal{A}(Y_1) - \mathcal{A}(Y_2)| \le \|Y_1 - Y_2\|_\infty .$$

Proof. Observe that there is a $\tilde{c} \in \mathbb{R}$ such that $\tilde{Y} \le \tilde{c}$. Whence $-\infty \le \mathcal{A}\left(\tilde{Y}\right) \le \mathcal{A}(\tilde{c})$ by monotonicity, thus further $-\infty < \mathcal{A}(c)$ for any real number c by translation equivariance.

Next find $c \in \mathbb{R}$ with $c \le Y$ a.s., therefore $-\infty < \mathcal{A}(c) \le \mathcal{A}(Y)$, so $\mathcal{A}$ is $\mathbb{R}-$valued for any $Y \in L^\infty$.

To observe continuity choose c_1 and c_2 such that $c_1 \le Y_1 - Y_2 \le c_2$ a.s.. From monotonicity and translation equivariance then follows that

$$\mathcal{A}(Y_1) = \mathcal{A}(Y_2 + Y_1 - Y_2) \le \mathcal{A}(Y_2 + c_2) = \mathcal{A}(Y_2) + c_2$$

and

$$\mathcal{A}(Y_2) = \mathcal{A}(Y_1 + Y_2 - Y_1) \le \mathcal{A}(Y_1 - c_1) = \mathcal{A}(Y_1) - c_1,$$

combined thus

$$c_1 \le \mathcal{A}(Y_1) - \mathcal{A}(Y_2) \le c_2.$$

Changing the role of Y_1 and Y_2 reveals the assertion. □

Similar statements hold for the general situation $1 \le p < \infty$ as well. They are more involved than the latter Lemma, we cite a precise statement from [SDR09, Proposition 6.7], its proof is built on Baire's lemma.

Theorem 5 ([SDR09, Proposition 6.7])**.** *Suppose that $\mathcal{A} : L^p \to \bar{\mathbb{R}}$ ($1 \le p < \infty$) is monotone, translation equivariant and concave, and further let $\{\mathcal{A} > -\infty\}$ have non-empty interior. Then $\mathcal{A}$ is finite valued and continuous.*

2.2. Subdifferential

Typical regularity properties of concave functionals include not only continuity, in many situations they are even subdifferentiable.

Definition. The *subdifferential* $\partial\mathcal{A}(Y)$ of a $\mathbb{R}-$valued function $\mathcal{A} : \mathcal{Y} \to \mathbb{R}$ is the collection of all *subgradients*,

$$\partial\mathcal{A}(Y) = \{Z^* \in \mathcal{Y}^* \colon \mathcal{A}(Y') - \mathcal{A}(Y) \le Z^*(Y' - Y) \text{ for all } Y' \in \mathcal{Y}\}.$$

$\mathcal{A}$ is called *sub-differentiable* at Y iff there is at least one subgradient, that is $\partial\mathcal{A}(Y)$ is non-empty.

For the Banach space $\mathcal{Y} = L^p(P)$ ($1 \le p < \infty$) one may choose $\mathcal{Y}^*$ the dual space of $\mathcal{Y}$. The inner product then is $Z^*(Y) = \mathbb{E}_P Z \cdot Y$ for some $Z \in L^{p'}(P)$ where p' is the conjugate exponent, $\frac{1}{p} + \frac{1}{p'} = 1$.

As for $\mathcal{Y} = L^\infty$ we follow the common practice and choose $\mathcal{Y}^* := L^1$ for the situation $p = \infty$, L^∞ thus is not paired with its dual space, but with the pre-dual instead.

Theorem 6 ([SDR09, Proposition 6.5])**.** *Let $\mathcal{A}: L^p \to \mathbb{R}$ ($p \leq \infty$) be a real-valued, monotone and concave acceptability functional. Then $\mathcal{A}$ is continuous and subdifferentiable on the entire L^p.*

The preceding Theorems elaborate on the situations where subgradients are available. The subgradients can be involved to derive a statement on continuity, as the following lemma reveals:

Lemma 7. *Let $\mathcal{A}$ be an acceptability functional on L^p, $1 \leq p < \infty$. Then*

$$\left|\mathcal{A}(Y) - \mathcal{A}(Y')\right| \leq \left\|Y - Y'\right\|_p \cdot \max\left\{\inf_{Z \in \partial\mathcal{A}(Y)} \|Z\|_{p'}, \inf_{Z \in \partial\mathcal{A}(Y')} \|Z\|_{p'}\right\},$$

in particular

$$\left|\mathcal{A}(Y) - \mathcal{A}(Y')\right| \leq \left\|Y - Y'\right\|_p \cdot \sup_{Z \in \partial A(Y), Z \in \partial A(Y')} \|Z\|_{p'}.$$

Proof. Let $Z \in \partial\mathcal{A}(Y)$ be chosen, that is

$$\mathcal{A}(Y') \leq \mathcal{A}(Y) + \mathbb{E}Z \cdot (Y' - Y).$$

By Hölder's inequality thus

$$\mathcal{A}(Y') - \mathcal{A}(Y) \leq \mathbb{E}Z \cdot (Y - Y') \leq \left\|Y - Y'\right\|_p \cdot \|Z\|_{p'}.$$

Interchanging the role of Y and Y' gives the first assertion, from which the other follows. □

2.3. Legendre-Fenchel Transformation

Associated with the subdifferential of a concave function $\mathcal{A}$ is the Legendre Fenchel transformation. The *conjugate function A of $\mathcal{A}$* is defined as

$$A(Z^*) := \inf_{Y \in \mathcal{Y}} Z^*(Y) - \mathcal{A}(Y)$$

on the space $\mathcal{Y}^*$, then, by the Fenchel-Moreau Theorem (cf. – for example – [RS06] or [Roc70]), $\mathcal{A}$ has the representation

$$\mathcal{A}(Y) = \inf_{Z^* \in \mathcal{Y}^*} Z^*(Y) - A(Z^*)$$

provided that $\mathcal{A}$ is upper semi-continuous.

We shall exploit this representation in various investigations below for functionals $\mathcal{A}$ defined on some $L^p(X, \mathcal{F}, P)$.

2.4. Wasserstein Metric

On a Polish space (X, d) consider the probability measures on its Borel sets. The collection of all probability measures, which satisfy for some – and thus any – $x_0 \in X$ the moment-like condition

$$\int_\Omega d(x_0, x)^r P(\mathrm{d}x) < \infty$$

is denoted by $\mathcal{P}_r(X; d)$.

On this space of probability measures define the function

$$\mathsf{d}_r(P_1, P_2; d) := \left(\inf_\pi \int_{X \times X} d(x_1, x_2)^r \pi(\mathrm{d}x_1, \mathrm{d}x_2)\right)^{\frac{1}{r}} \qquad (r \geq 1), \tag{6}$$

where the infimum is taken over all (bivariate) probability measures π on $X \times X$ with marginals P_1 and P_2, that is

$$\pi(A \times X) = P_1(A) \text{ and } \pi(X \times B) = P_2(B)$$

for all measurable sets $A \subset X$ and $B \subset X$ (in symbols $\pi_1 = P_1$, $\pi_2 = P_2$). We shall call such a measure π a *transport plan.*

d_r is called r^{th}-*Wasserstein distance.* It is well-defined, as for example the product measure $\pi := P_1 \otimes P_2$ has the required marginals and whence

$$\mathsf{d}_r\left(P_1, P_2; d\right)^r \leq \int_X \int_X d\left(x_1, x_2\right)^r P_1\left(\mathrm{d}x_1\right) P_2\left(\mathrm{d}x_2\right).$$

A very comprehensive and beautiful discussion and treatment of the distance d_r can be found in Villanis books ([Vil03] and [Vil09]), but we want to mention the books by Rachev and Rüschendorf as well, [RR98] and [GS02, PP11a].

We shall use the properties that the infimum in (6) is actually attained, and $\mathsf{d}_r(., .; d)$ turns out to be a metric on the space $\mathcal{P}_r(X; d)$, so particularly satisfies the triangle inequality

$$\mathsf{d}_r\left(P, P'; d\right) \leq \mathsf{d}_r\left(P, \tilde{P}; d\right) + \mathsf{d}_r\left(\tilde{P}, P'; d\right).$$

We are using the symbol d for the distance in the original space X, and $\mathsf{d}_r(.; d)$ to account for the distance on probabilities in $\mathcal{P}_r(X; d)$ induced by d. However, as no confusion may occur we will omit the distance d in the sequel and simply write $\mathsf{d}_r(P, P') = \mathsf{d}_r(P, P'; d)$ for the distance on $\mathcal{P}_r$ specified by d.

Remark 8. In honor of G. Monge[4] (cf. [Mon81]) and Leonid Kantorovich[5] (cf. [Kan42]) the distance d_r is sometimes called *Monge-Kantorovich distance* of order r, and d_2 is called quadratic Wasserstein distance as well. Moreover, the distance d_1 is also called Kantorovich-Rubinstein distance and sometimes denoted $\mathsf{d}_{KA} := \mathsf{d}_1$.

3. Continuity With Respect to Changing the Measure

Let $Y : X \to \mathbb{R}$ be a function, measurable with respect to $\mathcal{F}$. Then $Y \in \mathcal{Y} = L^p(X, \mathcal{F}, P)$, provided that $\int |X|^p \, \mathrm{d}P < \infty$, and the investigations in (the previous) Section 2 are concerned with the mapping

$$Y \mapsto \mathcal{A}(Y).$$

All acceptability functionals described above can be used now to compute $\mathcal{A}(Y)$, however, the result will depend on the probability measure P employed (cf. (1), (2), (3) and (4)). We shall make this evident by writing $\mathcal{A}_P(Y)$. The natural question, which arises in this context, is the question of *how* will $\mathcal{A}_P$ vary, whenever the measure P is changed? This is the topic of this section, we shall investigate the continuity properties of

$$P \mapsto \mathcal{A}_P(Y).$$

Theorem 9. *Let $Y : X \to \mathbb{R}$ be Hölder continuous with constant L_β, $|Y(x) - Y(x')| \leq L_\beta \cdot d(x, x')^\beta$ for some $\beta \leq 1$ and $\mathcal{A}$ law invariant. Assume that $Y \in L^p(P)$ and $Y \in L^p(P')$. Then*

$$\mathcal{A}_{P'}(Y) - \mathcal{A}_P(Y) \leq L_\beta(Y) \cdot \mathsf{d}_r(P, P') \cdot \inf_{A_P(Z) > -\infty} \|Z\|_{P, r_\beta} \tag{7}$$

and a fortiori

$$\left|\mathcal{A}_{P'}(Y) - \mathcal{A}_P(Y)\right| \leq L_\beta(Y) \cdot \mathsf{d}_r(P, P') \cdot \sup_{A_P(Z) > -\infty} \|Z\|_{P, r_\beta},$$

where $r_\beta \geq \frac{r}{r-\beta}$.

[4] Gaspard Monge (1746 - 1818) investigated how to efficiently construct dugouts.

[5] L. Kantorovich was awarded the price in Economic Sciences in Memory of Alfred Nobel in 1975.

Proof. Recall the representation

$$\mathcal{A}_P(Y)=\inf_{Z\in\mathcal{Y}^*}\mathbb{E}_P YZ-A_P(Z),$$

and, by weak* compactness, that there is an optimal Z_Y with $A_P(Z_Y)>-\infty$ such that infimum is attained, that is $\mathcal{A}_P(Y)=\mathbb{E}_P YZ_Y-A_P(Z_Y)$. For any Z' with $A_{P'}(Z')>-\infty$ thus

$$\begin{aligned}\mathcal{A}_{P'}(Y)-\mathcal{A}_P(Y)&\le\mathbb{E}_{P'}YZ'-A_{P'}(Z')-\mathbb{E}_P YZ_Y+A_P(Z_Y)\\&=\int Y(x')Z'(x')P'(\mathrm{d}x')-\int Y(x)Z_Y(x)P(\mathrm{d}x)-A_{P'}(Z')+A_P(Z_Y)\\&=\int Y(x')Z'(x')-Y(x)Z_Y(x)\,\pi(\mathrm{d}x,\mathrm{d}x')-A_{P'}(Z')+A_P(Z_Y)\end{aligned}\tag{8}$$

where π has marginals $\pi_1=P$ and $\pi_2=P'$. Taking the infimum over all these bivariate measures π and random variables Z' which have the same distribution as Z_Y under their respective measures (i.e. $P(Z_Y\le z)=P'(Z'\le z)$) one obtains

$$(8)\quad\le\quad\inf_{\pi,\,Z\sim Z_Y}\int Y(x')Z(x')-Y(x)Z_Y(x)\,\pi(\mathrm{d}x,\mathrm{d}x')-A_{P'}(Z)+A_P(Z_Y).$$

To ensure that such a random variable Z exists consider the cumulative distribution functions $G(z):=P(Z_Y\le z)$ and $G'(z):=P'(Z'\le z)$ with respective quantiles $G^{-1}(p)=\inf\{u\colon G(u)\ge p\}$ and $G'^{-1}(p)=\inf\{u\colon G'(u)\ge p\}$. Let $U'\colon X'\to\mathbb{R}$ be independent of Z' and uniformly distributed ($P'(U'\le u)=u$), and define the random variable $F(Z',U'):=(1-U')\cdot G'^{-}(Z')+U'\cdot G'(Z')$ where $G'^{-}(z)=P'(Z'<z)$. It is well-known (cf. [Fer67] for the so-called *generalized quantile transform*) that $F(Z',U')$ is uniformly distributed and moreover $Z'=G'^{-1}(F(Z',U'))$ $P'-$almost everywhere.

With these preparations define the random variable $Z:=G^{-1}(F(Z',U'))$. It holds that

$$\begin{aligned}P(Z_Y\le z)&=G(z)=P'(F(Z',U')\le G(z))\\&\le P'\left(G^{-1}(F(Z',U'))\le G^{-1}(G(z))\right)\\&\le P'\left(G^{-1}(F(Z',U'))\le z\right)=P'(Z\le z)\end{aligned}$$

because $F(Z',U')$ is uniform under P' and $G^{-1}(G(z))\le z$. Moreover

$$\begin{aligned}P'(Z\le z)&=P'\left(G^{-1}(F(Z',U'))\le z\right)\\&\le P'(F(Z',U')\le G(z))=G(z)=P(Z_Y\le z),\end{aligned}$$

such that $P(Z_Y\le z)=P'(Z\le z)$, that is Z and Z_Y have the same distribution given their respective measures and Z_Y can be replicated.

$\mathcal{A}$ is law invariant by assumption, whence so is A and thus $A_{P'}(Z)=A_P(Z_Y)$ (cf. [Sha11]) and

$$(8)\quad\le\quad\inf_{\pi,\,Z\sim Z_Y}\int Y(x')Z(x')-Y(x)Z_Y(x)\,\pi(\mathrm{d}x,\mathrm{d}x').$$

The infimum over Z in the latter expression is attained for some Z which is coupled in a antimonotone way with Y (cf. [Nel98]). By employing Hoeffding's Lemma ([Hoe40]) one may allow the random variable Z to be defined on the larger space $X\times X'$ without impacting the inf, the latter expression thus may be rewritten as

$$(8)\quad\le\quad\inf_{\pi,\,Z\sim Z_Y}\int Y(x')Z(x,x')-Y(x)Z_Y(x)\,\pi(\mathrm{d}x,\mathrm{d}x').$$

Whence, for the special choice $Z=Z_Y$,

$$\begin{aligned}(8)\quad&\le\quad\inf_{\pi}\int Y(x')Z_Y(x)-Y(x)Z_Y(x)\,\pi(\mathrm{d}x,\mathrm{d}x')\\&\le\quad\inf_{\pi}\int|Y(x')-Y(x)|\,Z_Y(x)\,\pi(\mathrm{d}x,\mathrm{d}x')\\&\le\quad\inf_{\pi}\int L_\beta\cdot d(x,x')^\beta\,Z_Y(x)\,\pi(\mathrm{d}x,\mathrm{d}x').\end{aligned}$$

For the relation $\frac{1}{\frac{r}{\beta}}+\frac{1}{\frac{r}{r-\beta}}=1$ one may apply Hölder's inequality to obtain

$$\begin{aligned}(8)\quad &\leq L_\beta\cdot\inf_\pi\left(\int d^r\mathrm{d}\pi\right)^{\frac{\beta}{r}}\cdot\left(\int Z_Y^{\frac{r}{r-\beta}}\mathrm{d}P\right)^{\frac{r-\beta}{r}}\\ &= L_\beta\cdot\mathsf{d}_r\left(P,P'\right)^\beta\cdot\left\|Z_Y\right\|_{P,\frac{r}{r-\beta}}.\end{aligned}$$

The other expression of the assertion follows by interchanging the role of P and P'. ☐

The following corollary quantifies the constants for the Average Value-at-Risk (cf. [PW09] for an initial statement in this direction) and for distortion acceptability functionals.

Corollary 10. *Let $Y\colon X\to\mathbb{R}$ be Hölder continuous with constant L_β, $\left|Y(x)-Y(x')\right|\leq L_\beta\cdot d(x,x')^\beta$ for some $\beta\leq 1$. Then*

$$\left|\mathrm{AV@R}_{\alpha,P}(Y)-\mathrm{AV@R}_{\alpha,P'}(Y)\right|\leq L_\beta(Y)\cdot\mathsf{d}_r\left(P,P'\right)\cdot\alpha^{-\frac{\beta}{r}}$$

and

$$\left|\mathcal{A}_h(Y)-\mathcal{A}_h(Y)\right|\leq L_\beta(Y)\cdot\mathsf{d}_r\left(P,P'\right)\cdot\|h\|_{r_\beta},$$

where $\mathcal{A}_h$ is a distortion acceptability functional and $\|h\|_r=\left(\int_0^1 h^r\right)^{\frac{1}{r}}$ is the norm for $h\in L^r([0,1],\lambda)$ on the standard space $([0,1],\lambda)$ with Lebesgue measure λ.

Proof. Recall the representations of the Average Value-at-Risk (cf. [PR07, Pfl06])

$$\begin{aligned}\mathrm{AV@R}_\alpha(Y) &= \inf\left\{\mathbb{E}YZ\colon 0\leq Z\leq\frac{1}{\alpha}\text{ and }\mathbb{E}Z=1\right\}\\ &= \inf\left\{\mathbb{E}Y h_\alpha(U): U\text{ uniformly distributed}\right\}\end{aligned}$$

where h_α is the distortion function $h_\alpha(u)=\begin{cases}\frac{1}{\alpha} & \text{if } u\leq\alpha\\ 1 & \text{if } u>\alpha\end{cases}$. The assertion follows from the other statement of the corollary, as

$$\|h_\alpha\|_{r_\beta}=\left(\int_0^\alpha\frac{1}{\alpha^{r_\beta}}\mathrm{d}P\right)^{\frac{1}{r_\beta}}=\left(\alpha^{1-r_\beta}\right)^{\frac{1}{r_\beta}}=\alpha^{\frac{1}{r_\beta}-1}=\alpha^{\frac{r-\beta}{r}-1}=\alpha^{-\frac{\beta}{r}}.$$

To prove the second statement we employ the relation

$$\mathcal{A}_h(Y)=\inf\left\{\mathbb{E}Yh(U): U\colon X\to[0,1]\ \text{ uniformly distributed}\right\}\tag{9}$$

from [PR07, Proposition 2.65], with Theorem 9 thus

$$\|h(U)\|_{r_\beta}^{r_\beta}=\int_X h(U)^{r_\beta}\,\mathrm{d}P=\int_0^1 h(u)^{r_\beta}\,\mathrm{d}u=\|h\|_{r_\beta}^{r_\beta}$$

for *any* uniformly distributed random variable $U\colon X\to[0,1]$ with law $P(U\leq u)=u$. ☐

Corollary 11. *Let $Y\colon X\to\mathbb{R}$ be Hölder continuous with constant L_β, $\left|Y(x)-Y(x')\right|\leq L_\beta\cdot d(x,x')^\beta$ for some $\beta\leq 1$ and $\mathcal{A}$ have the Kusuoka representation $\mathcal{A}(Y)=\inf_{m\in\mathscr{M}}\int\mathrm{AV@R}_p(Y)\,m(\mathrm{d}p)$. Then $\mathcal{A}$ is continuous with respect to the Wasserstein distance provided that*

$$K:=\sup_{m\in\mathscr{M}}\int_0^1\frac{1}{\alpha^{\frac{\beta}{r}}}m(\mathrm{d}\alpha)<\infty$$

is finite:

$$\left|\mathcal{A}_P(Y)-\mathcal{A}_{P'}(Y)\right|\leq\mathsf{d}_r\left(P,P'\right)\cdot L_\beta(Y)\cdot\sup_{m\in\mathscr{M}}\int_0^1\alpha^{-\frac{\beta}{r}}m(\mathrm{d}\alpha).$$

Remark 12. As $\int_0^1 p^{-\frac{\beta}{r}} m\left(\mathrm{d}p\right) \leq \int_0^1 \frac{1}{p} m\left(\mathrm{d}p\right)$ the statement particularly includes all distortion acceptability functionals due to Lemma 3.

Proof. For the straight forward proof choose $m_\varepsilon \in \mathscr{M}$ such that

$$\int_0^1 \mathrm{AV@R}_{\alpha,P'}\left(Y\right) m_\varepsilon\left(\mathrm{d}\alpha\right) < \mathcal{A}_{P'}\left(Y\right) + \varepsilon.$$

Then

$$\begin{aligned}
&\mathcal{A}_P\left(Y\right) - \mathcal{A}_{P'}\left(Y\right) = \\
&\quad \leq \int_0^1 \mathrm{AV@R}_{\alpha,P}\left(Y\right) m_\varepsilon\left(\mathrm{d}\alpha\right) - \int_0^1 \mathrm{AV@R}_{\alpha,P'}\left(Y\right) m_\varepsilon\left(\mathrm{d}\alpha\right) + \varepsilon \\
&\quad \leq \int_0^1 \mathrm{AV@R}_{\alpha,P}\left(Y\right) - \mathrm{AV@R}_{\alpha,P'}\left(Y\right) m_\varepsilon\left(\mathrm{d}\alpha\right) + \varepsilon \\
&\quad \leq L_\beta\left(Y\right) \cdot \mathsf{d}_r\left(P,P'\right)^\beta \cdot \int \frac{1}{\alpha^{\frac{\beta}{r}}} \mathrm{d}m^{G_\varepsilon}\left(\alpha\right) + \varepsilon \leq K \cdot L_\beta\left(Y\right) \cdot \mathsf{d}_r\left(P,P'\right)^\beta + \varepsilon.
\end{aligned}$$

Let $\varepsilon \to 0$ and interchange the role of P and P' to observe the desired assertion. ☐

4. Worst Measures

The problem investigated in the previous section can be restated as

$$\begin{array}{rl}
\underset{\left(\text{in } P'\right)}{\text{minimize}} & \mathcal{A}_{P'}\left(Y\right) \\
\text{subject to} & \mathsf{d}_r\left(P,P'\right) \leq K, \\
 & P' \in \mathcal{P}_r\left(X\right),
\end{array} \tag{10}$$

where the minimum (infimum) is among all probability measures P' whose Wasserstein-distance to P does not exceed K. A lower bound for the objective in (10) was found in Theorem 9, as

$$\mathcal{A}_{P'}\left(Y\right) \geq \mathcal{A}_P\left(Y\right) - L\left(Y\right) \cdot K \cdot \inf_{A_P(Z) > -\infty} \left\|Z\right\|_{r'}.$$

We shall drive the investigations further now and give some general situations for which this bound is sharp. It is possible in some situations to characterize the measure, for which problem (10) attains its minimal value. It will turn out that these measures have an interesting description as a transport map. Moreover, situations will occur where the bounds are not attained, and for some of them we will prove that no such bound exists in general.

Most of the results in this section are base on linear functionals Y. This is motivated from finance, as any typical optimal investment decision corresponds to a linear functional Y: Indeed, the return of the entire portfolio consisting of stocks s, each having random return x_s, is $\sum_s Y_s x_s =: Y\left(x\right)$ where Y_s corresponds with the total exposure in stock s. Above that the worst situations just occur for linear functionals, it is whence sufficient to restrict the considerations to linear functionals.

Before turning to the general situation (10) we start with the simpler problem

$$\begin{array}{rl}
\underset{\left(\text{in } P'\right)}{\text{minimize}} & \mathbb{E}_{P'} Y \\
\text{subject to} & \mathsf{d}_r\left(P,P'\right) \leq K, \\
 & P' \in \mathcal{P}_r\left(X\right)
\end{array} \tag{11}$$

to develop the strategy and the notion.

As above, let $(X, \mathcal{F}, P)$ denote a probability triple. We shall assume in addition that X is a linear space – in the simplest situation $X = \mathbb{R}^d$ – equipped with an appropriate norm function $\|\cdot\|$.

On this space there is the usual notion of a dual X^*, collecting all continuous, linear functionals on X. Recall that any linear functional $Y\colon X \to \mathbb{R}$ is a random variable itself, and $Y \in X^*$. The Lipschitz constant of Y, $L(Y) = \sup \frac{|Y(x)|}{\|x\|} = \|Y\|$ is the norm in the dual space.

Lemma 13. *Let $(X, \|\cdot\|)$ be a reflexive Banach space and Y a linear functional. Then, for all $1 \leq r < \infty$, the bound*

$$\mathbb{E}_P Y - K \cdot L(Y)$$

for (11) is sharp: There exists $x_Y \in X$, $\|x_Y\| = 1$, such that the minimizing measure in (11) is the push-forward (image measure of P)[6]

$$P^* := T_*(P) = P \circ T^{-1}$$

for the transport map (translation map) $T(x) := x - K \cdot x_Y$.

Proof. By Kantorovich's duality theorem (cf. [PP11b])

$$\begin{aligned} |\mathbb{E}_P Y - \mathbb{E}_{P'} Y| &\leq L(Y) \cdot \mathsf{d}_{KA}(P, P') \\ &= L(Y) \cdot \mathsf{d}_1(P, P') \leq L(Y) \cdot \mathsf{d}_r(P, P') \end{aligned} \tag{12}$$

for $r \geq 1$, establishing that $\mathbb{E}_{P'} Y \geq \mathbb{E}_P Y - L(Y) \cdot \mathsf{d}_r(P, P') \geq \mathbb{E}_P Y - L(Y) \cdot K$.

To observe that this bound is sharp recall that X and Y are linear. By the Hahn-Banach theorem $\|Y\| = x^{**}(Y)$ for some $x^{**} \in X^{**}$ with $\|x^{**}\| = 1$, and as X is reflexive one may identify x^{**} with some $x_Y \in X$, which satisfies

$$\|x_Y\| = 1 \text{ and } Y(x_Y) = \|Y\| = L(Y). \tag{13}$$

Define $T(x) := x - K \cdot x_Y$ and $P^* := T_*(P) = P \circ T^{-1}$ as above, and the push-forward transport plan

$$\pi := (\mathrm{id} \times T)_*(P),$$

for the mapping $(\mathrm{id} \times T)(x) := (x, T(x))$, that is $\pi(A \times B) = P\left(A \cap T^{-1}(B)\right)$.

The Wasserstein distance of P and P^* is bounded by K, because

$$\begin{aligned} \mathsf{d}_r(P, P^*)^r \leq \int d(x_1, x_2)^r \, \pi(\mathrm{d}x_1, \mathrm{d}x_2) &= \int d(x, T(x))^r \, P(\mathrm{d}x) \\ = \int \|K \cdot x_Y\|^r \, P(\mathrm{d}x) &= K^r \|x_Y\|^r = K^r. \end{aligned}$$

Given this measure P^* the objective of the primal function is

$$\begin{aligned} \mathbb{E}_{P^*} Y &= \int Y(x) \, P \circ T^{-1}(\mathrm{d}x) = \int Y(T(x)) \, P(\mathrm{d}x) \\ &= \int Y(x - K \cdot x_Y) \, P(\mathrm{d}x) = \mathbb{E}_P Y - K \cdot Y(x_Y) \\ &= \mathbb{E}_P Y - K \cdot L(Y), \end{aligned}$$

which is the minimum value we can achieve in view of (12). ☐

We shall now turn to the general situation.

[6] Villani rather uses the notation $T\#\mathbb{P} := T_*(\mathbb{P})$ for the push-forward measure, the notation $\mathbb{P}^T$ is in frequent use as well.

Theorem 14 (Optimal transport plan)**.** *Let Y be linear on a reflexive Banach space $(X, \|\cdot\|)$ and $\mathcal{A}$ law invariant.*

(i) *The problem*

$$\begin{array}{rl} \textit{minimize} & \\ \textit{(in } P'\textit{)} & \mathcal{A}_{P'}(Y) \\ \textit{subject to} & \mathsf{d}_r(P', P) \leq K \end{array} \tag{14}$$

has minimal value $\mathcal{A}_P(Y) - K \cdot L(Y) \cdot \min_{Z \in \partial\mathcal{A}_P(Y)} \|Z\|_{r'}$.

(ii) *For $1 < r < \infty$ there is a transport map such that*

$$P^* := T_*(P) = P \circ T^{-1}$$

minimizes (14).

Proof. Choose $x_Y \in X$ as in (13) and observe that $\partial A_P(Y)$ is weak* compact, one way hence select

$$Z_Y \in \operatorname{argmin}\left\{ \|Z\|_{\frac{r}{r-1}} : Z \in \partial\mathcal{A}_P(Y) \right\}. \tag{15}$$

Define the transport map

$$\begin{aligned} T(x) &:= x - K \cdot \left| \frac{Z_Y(x)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(x) \cdot x_Y \\ &= x - \frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}} \cdot |Z_Y(x)|^{\frac{r}{r-1}-2} \cdot Z_Y(x) \cdot x_Y, \end{aligned}$$

and again consider the transport plan

$$\pi := (\mathrm{id} \times T)_* (\mathbb{P})$$

with the marginals P and P^*. Observe that

$$\begin{aligned} \mathsf{d}_r(P, P^*)^r &\leq \int \|x - x'\|^r \, \pi(\mathrm{d}x, \mathrm{d}x') = \int \|x - T(x)\|^r \, P(\mathrm{d}x) \\ &= \int \left\| K \cdot \left| \frac{Z_Y(x)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot x_Y \right\|^r P(\mathrm{d}x) \\ &= \frac{K^r}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{r}{r-1}}} \cdot \int |Z_Y|^{\frac{r}{r-1}} \, \mathrm{d}P = K^r \end{aligned}$$

that is to say P^* has an accepted distance from P.

To observe that the transport map T is injective choose x_1 and x_2 and note that

$$\begin{aligned} T(x_1) - T(x_2) = & \\ = x_1 - x_2 + & \\ - K \cdot x_Y & \left(\left| \frac{Z_Y(x_1)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(x_1) - \left| \frac{Z_Y(x_2)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(x_2) \right). \end{aligned}$$

One may assume – without loss of generality – that $Z_Y(x_1) \leq Z_Y(x_2)$ (otherwise reverse x_1 and x_2) and distinguish the following two situations:

(i) If $Z_Y(x_1) = Z_Y(x_2)$, then $T(x_1) - T(x_2) = x_1 - x_2$ and T thus is injective on this subset.

(ii) If $Z_Y(x_1) < Z_Y(x_2)$, then $Y(x_1) \geq Y(x_2)$ a.s., because Y and Z_Y are coupled in an antimonotone way. In this situation

$$\begin{aligned}
&Y\left(T\left(x_1\right)-T\left(x_2\right)\right)= \\
&\quad= Y\left(x_1-x_2\right)+ \\
&\qquad -K\|Y\|\left(\left|\frac{Z_Y\left(x_1\right)}{\left\|Z_Y\right\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y\left(x_1\right)-\left|\frac{Z_Y\left(x_2\right)}{\left\|Z_Y\right\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y\left(x_2\right)\right) \\
&\quad> Y\left(x_1-x_2\right) \geq 0
\end{aligned}$$

because the map $x \mapsto \operatorname{sign}(x) \cdot |x|^{\frac{1}{r-1}}$ is increasing.

Whence, $T(x_1) \neq T(x_2)$ unless $x_1 = x_2$.

Define the random variable

$$Z_Y^T := \mathbb{E}\left[Z_Y \mid T\right]$$

by conditional expectation. Due to its definition Z_Y^T obeys the defining property

$$\int_{T^{-1}(B)} Z_Y \mathrm{d}P=\int_{T^{-1}(B)} \mathbb{E}\left[Z_Y \mid T\right] \circ T \mathrm{d}P=\int_B \mathbb{E}\left[Z_Y \mid T\right] \mathrm{d}T_*(P)=\int_B Z_Y^T \mathrm{d}P^* \tag{16}$$

for all measurable sets B (cf. [Wil91]). Notice, that

$$\int_{T^{-1}(B)} Z_Y \mathrm{d}P=\int_{T^{-1}(B)} Z_Y^T \circ T \mathrm{d}P=\int_B Z_Y^T \mathrm{d}P^*=\int_B Z_Y^T \mathrm{d}T_*(P)$$

by the change of variable formula again and for all measurable sets B, thus

$$Z_Y = Z_Y^T \circ T$$

$P-$almost everywhere, and

$$Z_Y^T = Z_Y \circ T^{-1}$$

P^*-almost everywhere as T is injective.

One deduces from (16) further that

$$\mathbb{E}_P Z_Y \cdot \mathbb{1}_{\{Z_Y \leq q\}}=\int_{T^{-1}T(\{Z_Y \leq q\})} Z_Y \mathrm{d}P=\int_{T\{Z_Y \leq q\}} Z_Y^T \mathrm{d}P^*=\mathbb{E}_{P^*} Z_Y^T \cdot \mathbb{1}_{\left\{Z_Y^T \leq q\right\}}$$

and whence

$$q-\frac{1}{\alpha} \mathbb{E}_P\left(q-Z_Y\right)^{+}=q-\frac{1}{\alpha} \mathbb{E}_{P^*}\left(q-Z_Y^T\right)^{+},$$

which is a well-known identity – cf. (2). Taking the maximum with respect to q, it will be attained for the same q at the left and at the right:

$$\begin{aligned}
G_{Z_Y}^{-1}(\alpha) &= \min \underset{q}{\operatorname{argmax}}\, q-\frac{1}{\alpha} \mathbb{E}_P\left(q-Z_Y\right)^{+} \\
&= \min \underset{q}{\operatorname{argmax}}\, q-\frac{1}{\alpha} \mathbb{E}_{P^*}\left(q-Z_Y^T\right)^{+}=G_{Z_Y^T}^{-1}(\alpha)
\end{aligned}$$

(a.e.) and so it follows that Z_Y and Z_Y^T have the same cumulative distribution function under their respective measures, $P(Z_Y \leq z) = P^*\left(Z_Y^T \leq z\right)$, so finally $A_{P^*}\left(Z_Y^T\right) = A_P(Z_Y)$.

As Z_Y is optimal by (15), $\mathcal{A}_P(Y) = \mathbb{E}_P Y \cdot Z_Y - A_P(Z_Y)$ and thus further

$$\begin{aligned}
\mathcal{A}_{P^*}(Y)-\mathcal{A}_P(Y) &= \\
&\leq \mathbb{E}_{P^*} Y \cdot Z_Y^T-A_{P^*}\left(Z_Y^T\right)-\mathbb{E}_P Y \cdot Z_Y+A_P\left(Z_Y\right) \\
&= \mathbb{E}_{P^*} Y \cdot Z_Y^T-\mathbb{E}_P Y \cdot Z_Y \\
&= \mathbb{E}_P(Y \circ T) \cdot Z_Y-\mathbb{E}_P Y \cdot Z_Y=\mathbb{E}_P Y(T-\mathrm{id}) \cdot Z_Y,
\end{aligned}$$

by linearity of Y. Using $Y(x_Y)=\|Y\|=L(Y)$ one finds further that

$$\begin{aligned}
&\mathcal{A}_{P^*}(Y)-\mathcal{A}_P(Y)\\
&\quad\leq \mathbb{E}_P Y\left(-K\cdot\left|\frac{Z_Y(x)}{\|Z_Y\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}}\operatorname{sign} Z_Y(x)\cdot x_Y\right)\cdot Z_Y\\
&\quad=-\frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}}\|Y\|\cdot\mathbb{E}_P|Z_Y|^{\frac{1}{r-1}}\cdot|Z_Y|=-\frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}}\|Y\|\cdot\mathbb{E}_P|Z_Y|^{\frac{r}{r-1}}\\
&\quad=-\frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}}\cdot L(Y)\cdot\|Z_Y\|_{\frac{r}{r-1}}^{\frac{r}{r-1}}\leq -K\cdot L(Y)\cdot\min_{Z\in\partial\mathcal{A}(Y)}\|Z\|_{\frac{r}{r-1}},
\end{aligned}$$

whence

$$\mathcal{A}_{P^*}(Y)-\mathcal{A}_P(Y)\leq -K\cdot L(Y)\cdot\min_{Z\in\partial\mathcal{A}(Y)}\|Z\|_{\frac{r}{r-1}}.$$

In view of (7) this is smallest difference achievable. □

The situation is a bit more involved for the Kantorovich distance, $r=1$.

Theorem 15. *Problem (14) has an optimal solution provided that $P(Z_Y=\|Z_Y\|)>0$, where Z_Y is as in (15). The corresponding transport map is*

$$T(x):=x-K\cdot\frac{\mathbb{1}_{\{|Z_Y|=\|Z_Y\|_\infty\}}(x)}{\mathbb{P}(|Z_Y|=\|Z_Y\|_\infty)}\cdot\operatorname{sign} Z_Y(x)\cdot x_Y.$$

Proof. For the Kantorovich distance ($r=1$) the proof needs a slight modification, it may read as follows:

$$\begin{aligned}
\mathsf{d}_{KA}(P,P^*)&\leq\int\|x-T(x)\|\,P(\mathrm{d}x)\\
&=\int\left\|K\cdot\mathbb{1}_{\{|Z_Y|=\|Z_Y\|_\infty\}}(x)\frac{\operatorname{sign} Z_Y(x)}{\mathbb{P}(|Z_Y|=\|Z_Y\|_\infty)}\cdot x_Y\right\|P(\mathrm{d}x)\\
&=K\cdot\int\frac{\mathbb{1}_{\{|Z_Y|=\|Z_Y\|_\infty\}}(x)}{P(|Z_Y|=\|Z_Y\|_\infty)}P(\mathrm{d}x)=K
\end{aligned}$$

On the other side,

$$\begin{aligned}
&\mathcal{A}_{P^*}(Y)-\mathcal{A}_P(Y)=\\
&\quad=\mathbb{E}_P Y\left(-K\cdot\mathbb{1}_{\{|Z_Y|=\|Z_Y\|_\infty\}}(x)\frac{\operatorname{sign} Z_Y(x)}{P(|Z_Y|=\|Z_Y\|_\infty)}\cdot x_Y\right)\cdot Z_Y\\
&\quad=-K\cdot\|Y\|\cdot\int\frac{\mathbb{1}_{\{|Z_Y|=\|Z_Y\|_\infty\}}|Z_Y|}{P(|Z_Y|=\|Z_Y\|_\infty)}\mathrm{d}P\\
&\quad=-K\cdot L(Y)\cdot\|Z_Y\|_\infty\leq -K\cdot L(Y)\cdot\min_{Z_Y\in\partial\mathcal{A}(Y)}\|Z_Y\|_\infty,
\end{aligned}$$

which establishes the result in this particular case. □

5. Continuity for the Distorted Functional $\mathcal{A}_h$

As was elaborated in Theorem 14 the optimal transport map may always be given provided that $r>1$. As for $r=1$ the additional requirement

$$P(|Z_Y|=\|Z_Y\|_\infty)>0$$

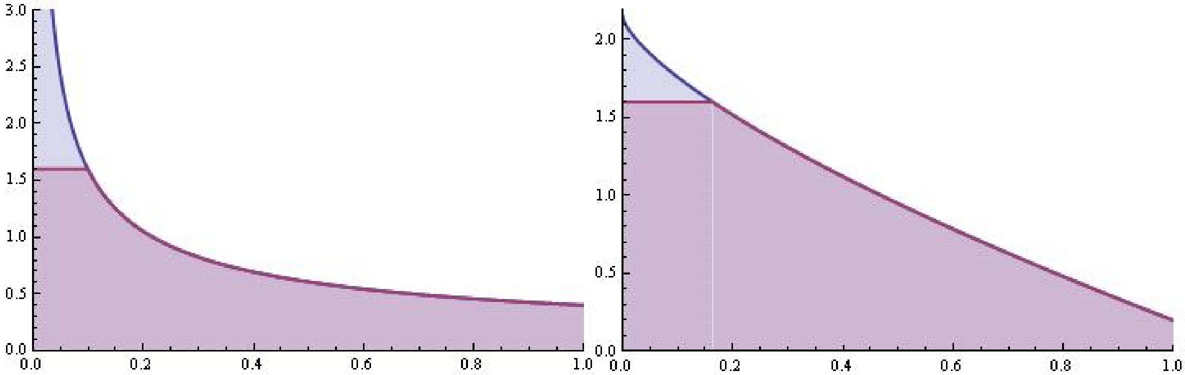

Figure 1: Exemplary shape for a bounded, and an unbounded distortion h. The area under both charts is one.

is needed to guarantee existence of the furthermost measure. We shall continue the discussion at this point and elaborate the continuity properties for the Kantorovich distance further.

Recall that $Z_Y = h(U)$ (cf. (9)) for some uniform distribution U for the distorted acceptability functional. The latter condition $P(|Z_Y| = \|Z_Y\|_\infty) > 0$ thus holds iff

$$\lambda(h = \|h\|_\infty) > 0,$$

where λ is the Lebesgue measure on $[0, 1]$: this particularly holds for AV@R's distortion function $h_a = \frac{1}{\alpha}\mathbb{1}_{[0,\alpha]}$, as

$$\lambda(h_\alpha = \|h_\alpha\|_\infty) = \alpha > 0,$$

and the optimal measure can be given in this situation for any $r \geq 1$ as indicated.

We shall now further discuss the properties in the degenerate situation where $\lambda(h = \|h\|_\infty) = 0$, in particular where h is

(i) unbounded (cf. Figure 1, left graph), and

(ii) h is bounded, but not flat at its top (Figure 1, right graph).

It will turn out that the first problem is pretty easy, whereas the second involves tough mathematical results.

5.1. Unbounded Distortions

Theorem 16. *Suppose that h is not bounded and Y is linear on a linear space. Then the problem*

$$\begin{array}{rl} \textit{minimize} & \\ \textit{(in } P'\textit{)} & \mathcal{A}_{h;P'}(Y) \\ \textit{subject to} & \mathsf{d}_{KA}(P, P') \leq K \end{array}$$

is not bounded neither, i. e. the objective is $-\infty$.

Proof. Consider the measures

$$P'_n := (T_n)_*(P),$$

for the transport plans

$$T_n(x) := x - K \cdot \frac{\mathbb{1}_{\{|Z_Y| \geq n\}}(x)}{P(|Z_Y| \geq n)} \cdot \operatorname{sign} Z_Y(x) \cdot x_Y$$

which cut the (possibly sub-optimal) dual variable $Z_Y = h(U)$.

As above $\mathsf{d}_{KA}(P'_n, P) = K$, but

$$\mathcal{A}_{h,P'_n}(Y) - \mathcal{A}_{h,P}(Y) \quad \leq \quad -K \cdot L(Y) \cdot \int |Z_Y| \cdot \frac{\mathbb{1}_{\{|Z_Y| \geq n\}}(x)}{P(|Z_Y| \geq n)} P(\mathrm{d}x) \leq -K \cdot L(Y) \cdot n$$

for any $n \in \mathbb{N}$, the problem thus does not allow a bounded (real) solution. $\square$

5.2. Bounded Distortions

Theorem 17. *Let Y be a (continuous) linear functional on $(\mathbb{R}^m, \|.\|)$. Moreover assume that h is bounded, but $\lambda(|h| = \|h\|_\infty) = 0$ (cf. Figure 1). Then the problem*

$$\begin{array}{rl} \text{minimize} & \\ \text{(in } P'\text{)} & \mathcal{A}_{h;P'}(Y) \\ \text{subject to} & \mathsf{d}_{KA}(P', P) \le K \end{array} \tag{17}$$

is bounded, but there does not *exist a measure P' with $\mathsf{d}_{KA}(P, P') \le K$ attaining the minimum in (17), that is to say the respective* argmin*-set is empty.*

Remark 18. Notice, that the latter statement holds true on finite dimensional $(\mathbb{R}^d, \|.\|)$, so there is no chance on infinite dimensional spaces neither to find a minimizing measure.

Proof. Define the set

$$C := \operatorname{argmin} \{\mathcal{A}_{h;P'}(Y) : \mathsf{d}_{KA}(P, P') \le K\},$$

which is the argmin-set, consisting of all measures minimizing the problem (cf. (10)) in consideration.

In order to prove the statement by contradiction suppose that C were not empty. As the optimal value is known precisely the minimum value of the problem may be written as

$$C = \{P' : \mathcal{A}_{h;P'}(Y) - \mathcal{A}_{h;P}(Y) = -K \cdot L(Y) \cdot \|h\|_\infty, \mathsf{d}_{KA}(P, P') \le K\}.$$

Further one may write

$$C = \bigcap_{n>1} C_n,$$

where the sets C_n originate from the relaxed problem

$$C_n = \left\{P' : \mathcal{A}_{h;P'}(Y) - \mathcal{A}_{h;P}(Y) \le -K \cdot L(Y) \cdot \left(\|h\|_\infty - \frac{1}{n}\right), \mathsf{d}_{KA}(P, P') \le K\right\};$$

those sets C_n are certainly non-empty.

Consider the measures

$$P'_n := (T_n)_* (P),$$

defined via the transport maps

$$T_n(x) := x - K \cdot \frac{\mathbb{1}_{\left\{|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right\}}(x)}{P\left(|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right)} \cdot \operatorname{sign} Z_Y(x) \cdot x_Y$$

by appropriately cutting the dual variable Z_Y at its top.

By the same computation as above they satisfy $\mathsf{d}_{KA}(P'_n, P) = K$ by construction, and

$$\begin{aligned} \mathcal{A}_{h,P'_n}(Y) - \mathcal{A}_{h,P}(Y) &\le -K \cdot L(Y) \cdot \int |Z_Y| \cdot \frac{\mathbb{1}_{\left\{|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right\}}(x)}{P\left(|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right)} P(\mathrm{d}x) \\ &\le -K \cdot L(Y) \cdot \left(\|Z_Y\|_\infty - \frac{1}{n}\right), \end{aligned}$$

and thus $P'_n \in C_n$.

As $(\mathbb{R}^d, \|.\|)$ is locally compact the space of continuous functions vanishing at infinity, $C_0(\mathbb{R}^d, \|.\|)$, is a Banach space, and Riesz' theorem identifies its dual with the space of regular Borel measures (cf. [Woj91]).

The probability measures P'_n may be considered themselves as elements of this dual via the natural setting

$$\begin{aligned} P'_n : C_0(\mathbb{R}^m) &\to \mathbb{R} \\ \varphi &\mapsto \int \varphi \mathrm{d}P'_n, \end{aligned}$$

but moreover

$$|P'_n(\varphi)| \le \int \|\varphi\|_\infty \, \mathrm{d}P'_n = \|\varphi\|_\infty$$

for any function $\varphi \in C_0\left(\mathbb{R}^d, \|.\|\right)$, and thus $\|P'_n\| \le 1$: That is to say all those measures P'_n are within the unit ball $B_1(0)$ of the dual of $C_0\left(\mathbb{R}^d, \|.\|\right)$.

Alaoglu's theorem states that the closed unit ball $B_1(0)$ in the dual is weakly* compact, thus there is an accumulation point $\tilde{P}' \in B_1(0)$ such that

$$P'_{n_k} \to \tilde{P}'$$

in the weak* topology for some sub-sequence $(n_k)_k$. Again by Riesz' theorem $\tilde{P}'$ has a representation as a measure, although not necessarily as a probability measure.

We shall prove next that C is convex. This holds true, because

(i) the distance d_{KA} is convex for the situation $r = 1$, that is [7]

$$\mathsf{d}_{KA}(P, (1-\lambda)P_0 + \lambda P_1) \le (1-\lambda)\,\mathsf{d}_{KA}(P, P_0) + \lambda \mathsf{d}_{KA}(P, P_1). \tag{18}$$

Indeed, let π_0 (π_1, resp.) have marginals P and P_0 (P and P_1, resp.), then $\pi_\lambda := (1-\lambda)\pi_0 + \lambda\pi_1$ has marginals P and $P_\lambda := (1-\lambda)P_0 + \lambda P_1$, such that (18) is immediate.

(ii) $P \mapsto \mathcal{A}_{h,P}(Y)$ is convex: to accept this consider $P_0 \in C$, $P_1 \in C$ and observe that the distribution functions

$$\begin{aligned} G_{Y,\lambda}(z) &:= P_\lambda(Y \le z) = (1-\lambda)P_0(Y \le z) + \lambda P_1(Y \le z) \\ &= (1-\lambda)G_{Y,0}(z) + \lambda G_{Y,1}(z) \end{aligned}$$

are convex-combinations. Whence

$$\begin{aligned} \mathcal{A}_{h,P_\lambda}(Y) &= \int G_{Y,\lambda}(z)\,h(z)\,\mathrm{d}z = \int ((1-\lambda)G_{Y,0}(z) + \lambda G_{Y,1}(z))\,h(z)\,\mathrm{d}z \\ &= (1-\lambda)\int G_{Y,0}(z)\,h(z)\,\mathrm{d}z + \lambda \int G_{Y,1}(z)\,h(z)\,\mathrm{d}z \\ &= (1-\lambda)\,\mathcal{A}_{h,P_0}(Y) + \lambda \mathcal{A}_{h,P_0}(Y) \end{aligned}$$

is a convex combination as well.

So C is convex. By Mazur's theorem the norm-closure and its weak* closure coincide for convex sets,

$$\tilde{P}' \in \overline{C}^{\text{weak*}} = \overline{C}^{\|\cdot\|},$$

we thus deduce in particular that

$$\left\|\tilde{P}'\right\| = 1,$$

and the limiting measure $\tilde{P}'$ thus is a probability measure.

Now define the increasing sets $X_n := \left\{|Z_Y| \le \|Z_Y\|_\infty - \frac{1}{n}\right\} \subset X$. Observe that

$$P'_n\left(\bigcup_n X_n\right) \ge P'_n(X_n) \ge \mathbb{P}\left(|Z_Y| \le \|Z_Y\|_\infty - \frac{1}{n}\right) = \lambda\left(h \le \|h\|_\infty - \frac{1}{n}\right) \longrightarrow 1$$

due to our assumptions, and particularly because $Z_Y = h(U)$. Whence $\tilde{P}'_n\left(\bigcup_n X_n\right) = 1$, and consequently $\tilde{P}'\left(\bigcup_n X_n\right) = 1$, because

$$P'_{n_k} \to \tilde{P}'$$

[7] (18) does not hold whenever $r > 1$.

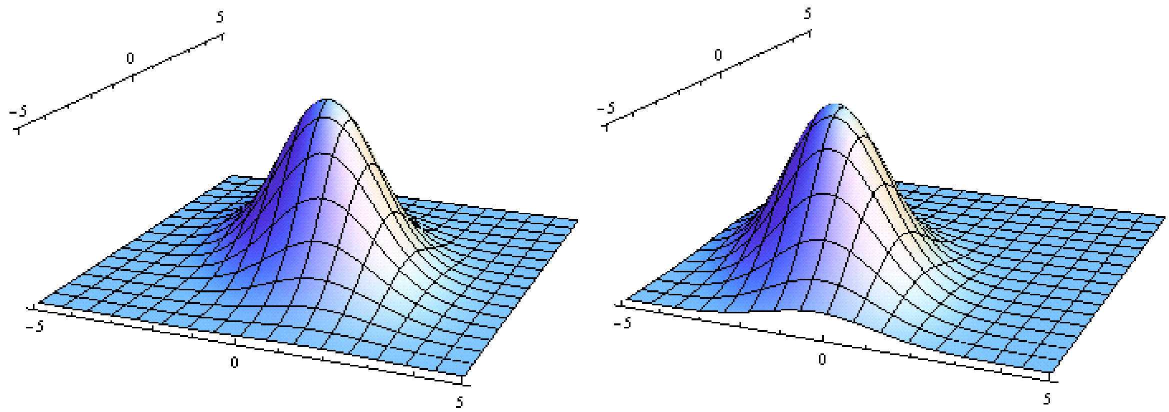

Figure 2: For the expectation, the worst measure is a simple translate. Here, $Y(x_1, x_2) = x_1 + x_2$ and whence the direction of the translation is $-x_Y = -\frac{1}{\sqrt{2}}(1,1)$ for the Euclidean distance.

by Portmanteau's Lemma (cf. [vdV98]). By construction (recall the definition of the transport map T_n), P'_n and P coincide on any X_n, so $\tilde{P}'$ and P coincide on every set $A \subseteq \bigcup_n X_n$. This, however, means $\tilde{P}' = P$, because

$$\tilde{P}'\left(\bigcup_n X_n\right) = P\left(\bigcup_n X_n\right) = 1.$$

This is a contradiction, because the measure P certainly is not optimal for the problem (17).

Whence, C is the empty set,

$$C = \emptyset,$$

and there is no optimal measure P' for the problem (17). □

6. Illustration of optimal Transport Maps

6.1. Expectation

The expectation is the simplest acceptability functional, $\mathcal{A} := \mathbb{E}$, by Theorem 14 the transport map for which

$$\mathcal{A}_{P'}(Y) = \mathcal{A}_P(Y) - K \cdot L(Y)$$

reduces to the simple translation $T(x) := x - K \cdot x_Y$ in this situation for all $1 \leq r < \infty$, which is exemplary depicted in Figure 2.

6.2. Distortions

For distortions we have elaborated that Z_Y is coupled in an antimonotone way with Y and moreover $Z_Y = h(U)$: We thus can give the dual variable as $Z_Y = h(G_Y(Y))$[8], and the transport map rewrites

$$T(x) = x - K \cdot \left|\frac{h(G_Y(Y(x)))}{\|h\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}} \cdot x_Y$$

for non-negative distortion functions h. This enables to illustrate the geometry by plotting some densities, which we want to do here in providing some examples: Two distortions of the same bivariate

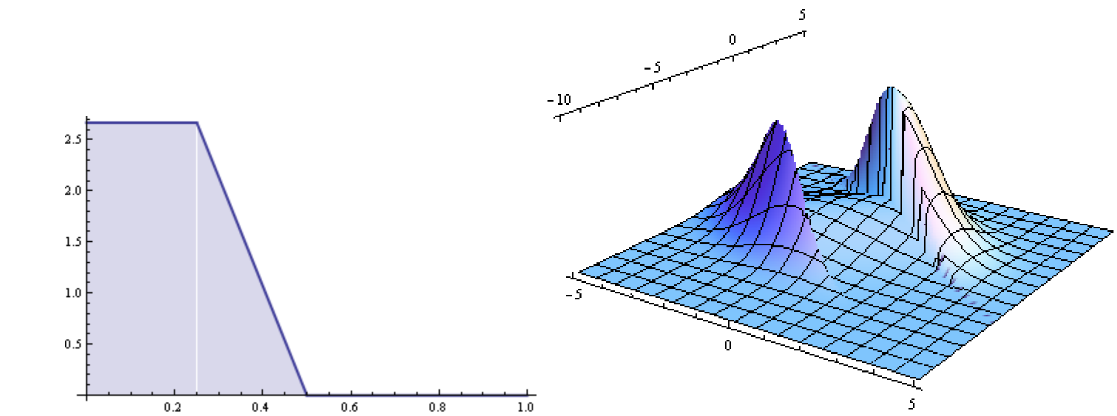

Figure 3: Distortion function h (left chart) and distorted probability measure (right chart) of the same measure as in Figure 2: 50 % stay at the same place, 25% of the mass is simply being shifted in direction $-x_Y$, and the remaining 25% are brutally distorted in between (Y as in the previous example).

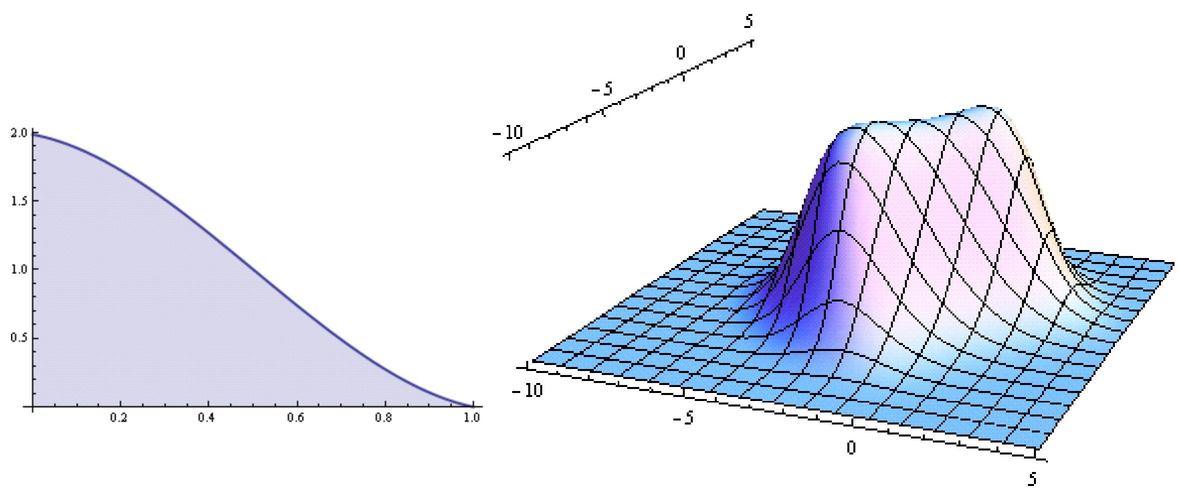

Figure 4: Resulting probability distribution (right chart) by applying the distortion function indicated (left chart) to the initial distribution from Figure 2.

distribution as in Figure 2 are depicted in Figure 3 and Figure 4, each with its corresponding distortion function h.[9]

[8]Recall that h is *de*creasing.

[9]Let P denote a bivariate normal probability measure with mean μ and covariance matrix Σ ($P \sim \mathcal{N}(\mu, \Sigma)$) and Y a linear functional of the form $Y(x) = Y^\top x = \sum_i Y_i x_i$, then $Y_* P \sim \mathcal{N}\left(Y^\top \mu, Y^\top \Sigma Y\right)$, that is $Y_*(P) \sim \mathcal{N}\left(\sum_i Y_i \mu_i, \sum_{i,j} Y_i \Sigma_{i,j} Y_j\right)$; whence, $G_Y(y) = \frac{1}{\sqrt{2\pi Y^\top \Sigma Y}} \int_{-\infty}^{y} e^{-\frac{1}{2}\frac{\left(x - Y^\top \mu\right)^2}{Y^\top \Sigma Y}} \mathrm{d}x$ and $G_Y(Y(x)) = \frac{1}{\sqrt{2\pi Y^\top \Sigma Y}} \int_{-\infty}^{Y^\top x} e^{-\frac{1}{2}\frac{\left(x' - Y^\top \mu\right)^2}{Y^\top \Sigma Y}} \mathrm{d}x'$ is a $\mathbb{R}$-valued random variable.

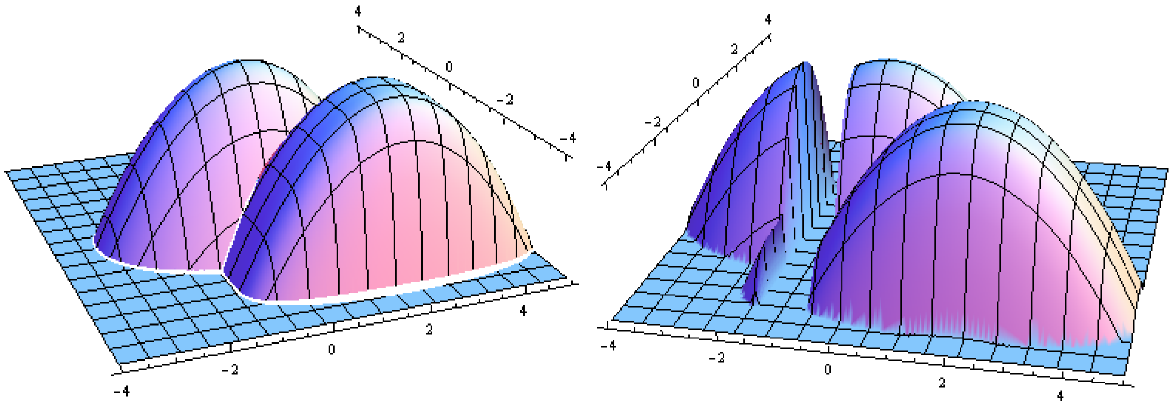

Figure 5: Initial (left) and split (right) probability measure, as it is worst with respect to the Average Value-at-Risk. Displayed from different perspectives.

6.3. The Average Value-at-Risk

As for AV@R$_\alpha$ the optimal dual variable basically is $Z_Y = \mathbb{1}_{\left\{Y \le G_Y^{-1}(\alpha)\right\}}$. The transport map, for all $1 \le r < \infty$, is

$$T(x) = x - \frac{K}{\alpha} \cdot x_Y \cdot \mathbb{1}_{\left\{Y < G_Y^{-1}(\alpha)\right\}}(x),$$

which again includes the expectation for $\alpha = 1$.

This transport map splits the sample space X according the α-quantile: Those samples, which *do not* contribute to the computation of AV@R$_\alpha$ (which have quantile $\left(\left\{Y > G_Y^{-1}(\alpha)\right\}\right)$, are left unchanged on their place, while all other samples, which *do* contribute to the AV@R$_\alpha$ $\left(\left\{Y < G_Y^{-1}(\alpha)\right\}\right)$, are being simply worsened by shifting them the distance $\frac{K}{\alpha}$; moreover, all of them are being shifted

- in *parallel*
- in the *same direction* $-x_Y$ and
- the *same distance* $\frac{K}{\alpha}$,

as exemplary indicated in Figure 5.

7. Summary

In the present paper we investigate the impact of probability measures on the evaluation of acceptability functionals. It is proved that the Wasserstein distance is a very useful notion of distance in the present context, as it allows precise bounds on potential evaluations.

The measures within a given Wasserstein-ball of radius K, which have the highest impact on the evaluation of an acceptability functional, can be described in some intuitive way, in particular for the Average Value-at-Risk.

This is a positive result as regards the adoption of acceptability in a wide range of applications in finance (cf. [PPW11]) in energy and far more areas, among them all considerations on optimization, robust optimization in particular.

8. Acknowledgment

We wish to thank Georg Ch. Pflug for his continual advise for the present paper.

And moreover the referees for their constructive criticism.

The plots have been produced by use of Mathematica.

9. Bibliography

[ADEH99] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999. 1

[Fer67] Thomas S. Ferguson. *Mathematical Statistics. A Decision Theoretic Approach.* Academic Press, New York, 1967. 7

[GS02] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002. 6

[Hoe40] Wassilij Hoeffding. Maßstabinvariante Korrelationstheorie. *Schriften Math. Inst. Univ. Berlin*, 5:181–233, 1940. 7

[JST06] Elyès Jouini, Walter Schachermayer, and Nizar Touzi. Law invariant risk measures have the Fatou property. *Advances in mathematical economics.*, 9:49–71, 2006. 2

[Kan42] Leonid Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942. 6

[Kus01] Shigeo Kusuoka. On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95, 2001. 2

[Mon81] Gaspard Monge. Mémoire sue la théorie des déblais et de remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704, 1781. 6

[Nel98] Roger B. Nelsen. *An Introduction to Copulas.* Number 139 in Lecture Notes in Statistics. Springer, 1998. 7

[Pfl06] Georg Ch. Pflug. On distortion functionals. *Statistics and Decisions*, 24:45–60, 2006. 2, 8

[PP11a] Georg Ch. Pflug and Alois Pichler. *Approximations for Probability Distributions and Stochastic Optimization Problems*, volume 163 of *International Series in Operations Research & Management Science*, chapter 15, pages 343–387. Springer New York, 2011. 6

[PP11b] Georg Ch. Pflug and Alois Pichler. A distance for multistage stochastic optimization models. *SIAM Journal on Optimization*, 2011. 10

[PPW11] Georg Ch. Pflug, Alois Pichler, and David Wozabal. The 1/ n investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 2011. submitted. 19

[PR07] Georg Ch. Pflug and Werner Römisch. *Modeling, Measuring and Managing Risk.* World Scientific, August 2007. 8

[PW09] Georg Ch. Pflug and David Wozabal. *Quantitative Fund Management.* Financial Mathematics Series. Chapman & Hall/ CRC, 2009. 8

[Roc70] R. Tyrrell Rockafellar. *Convex Analysis.* Princeton University Press, 1970. 5

[RR98] Svetlozar T. Rachev and Ludger Rüschendorf. *Mass Transportation Problems Vol. I: Theory, Vol. II: Applications.* Probability and its applications. Springer-Verlag, New York, 1998. 6

[RS06] Andrzej Ruszczyński and Alexander Shapiro. Optimization of convex risk functions. *Mathematics of operations research*, 31:433–452, 2006. 5

[RU00] R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000. 2

[SDR09] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming*. MPS-SIAM Series on Optimization, 2009. 4, 5

[Sha11] Alexander Shapiro. On Kusuoka representations of law invariant risk measures. 2011. 7

[vdV98] Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. 17

[Vil03] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003. 6

[Vil09] Cédric Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer, 2009. 6

[Wil91] David Williams. *Probability with Martingales*. Cambridge University Press, 1991. 12

[Woj91] Przemysław Wojtaszczyk. *Banach Spaces for Analysts*. Cambridge University Press, Cambridge, 1991. 15